An Energy-aware and Self-adaptive Approach for AI-based Applications in Smart Cities

Alessandro Tundo[1,2], Marco Mobilio[1], Shashikant Ilager[2], Ivona Brandić[2], Ezio Bartocci[2], Leonardo Mariani[1]

[1]University of Milano-Bicocca, Milan, Italy

[2]Vienna University of Technology, Vienna, Austria

This poster has been designed using images from Flaticon.com

9 INDUSTRY, INNOVATION AND INFRASTRUCTURE

11 SUSTAINABLE CITIES AND COMMUNITIES

The need for adaptable and controllable AI-based applications on the Edge
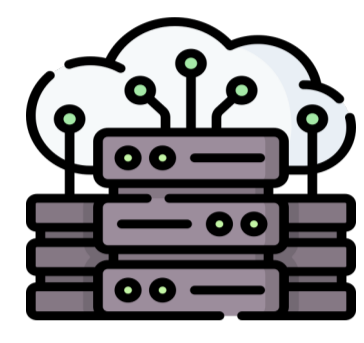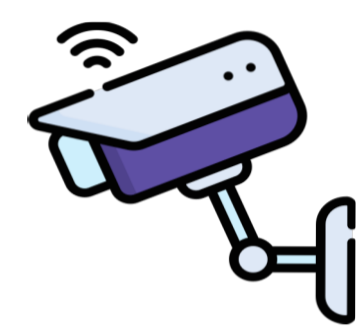
Many Scenarios

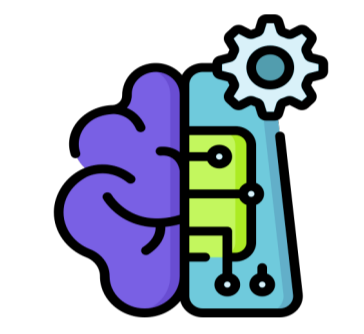*Night | Few pedestrians*

Day | Crowd

Complex Run-Time Environment

- Large number of **heterogeneous nodes**
- **Resource constrained** environment
- Potentially **unreliable power supply**

- **Interaction** with / **configuration** of **Cyber-Physical Systems**
- Large number of **heterogeneous IoT devices**

- **Configuration** of **AI model** parameters
- Usage of **resource-intensive accelerators**

Conflicting Objectives

1 **Being accurate**

2 **Consuming as little energy as possible**
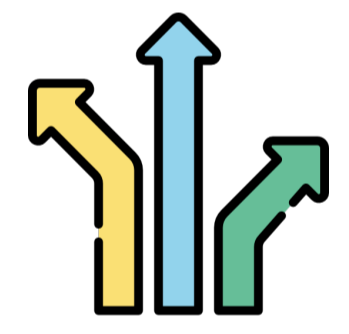
3 **Fast data processing**

...and potentially more!

Main challenges
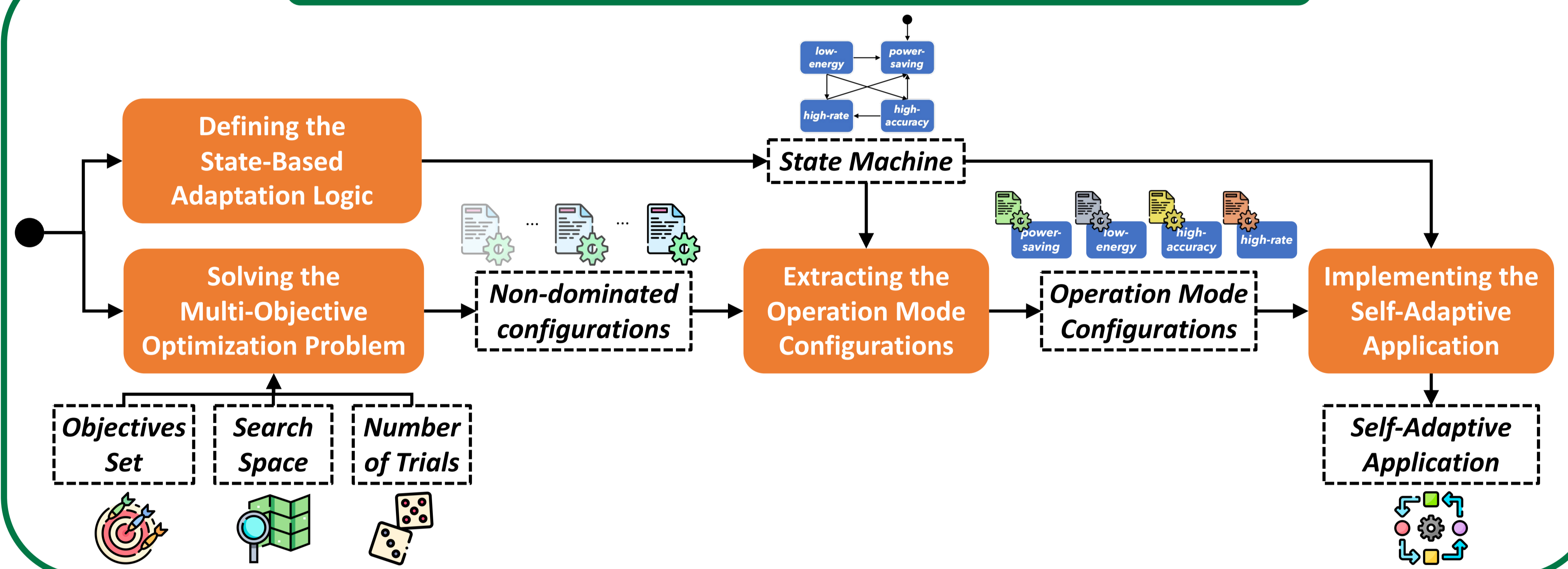
- **Lack** of a **globally optimized configuration**

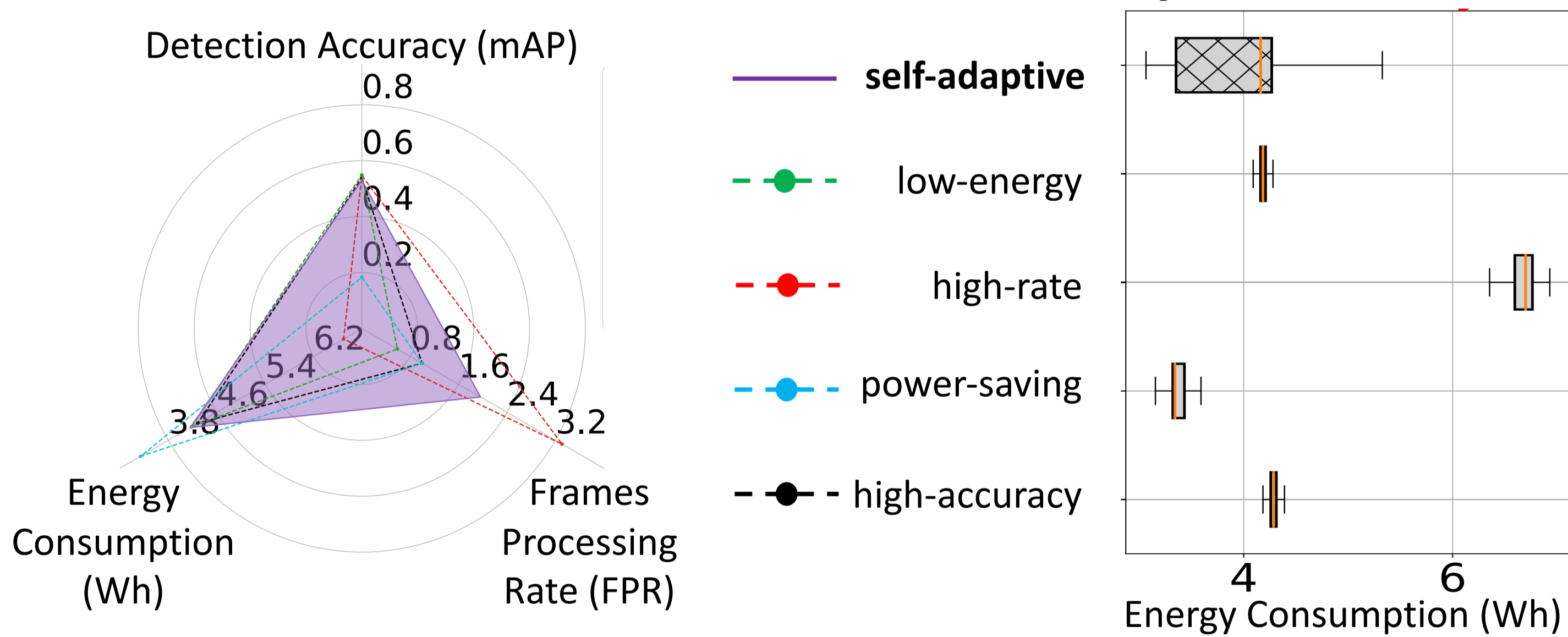- **Huge cost** for **configuration space exploration**

- **Different configurations** for **different** run-time **scenarios**

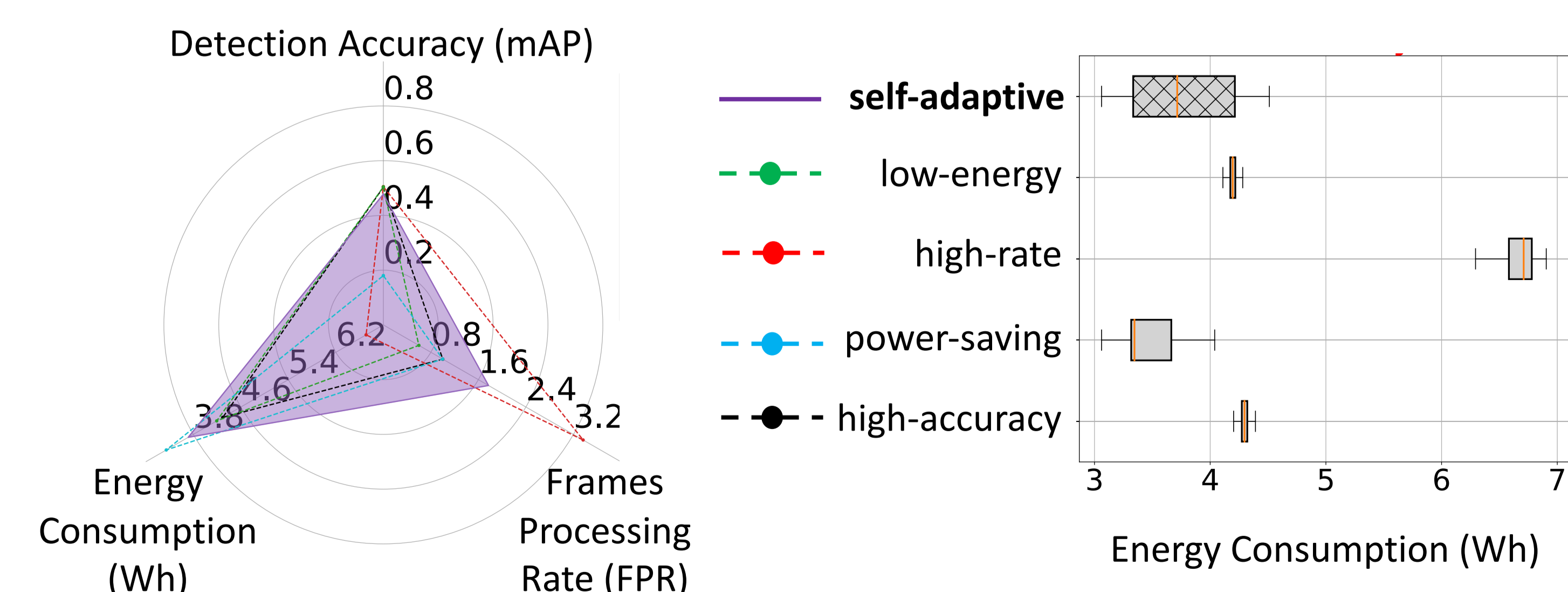## The proposed energy-aware approach

Experimental results

Pedestrian Detection Weekdays

Pedestrian Detection Weekends

Conclusions

- **Meta-heuristic** search procedure as **effective** as the **near-exhaustive** despite an empirical exploration of **only 10%** of the **search space**

- **Self-adaptive** solution **employs** more **accurate** operation **modes when workload** is **higher**, using **less accurate** operation **modes when** the **workload** is **less demanding**

- **Self-adaptive** solution **consumes energy only when** it is **worth**

References

Jiang, C., Fan, T., Gao, H., Shi, W., Liu, L., Cérin, C., & Wan, J. (2020). Energy aware edge computing: A survey. *Computer Communications*, 151, 556-580.

Panerati, J., Sciuto, D., & Beltrame, G. (2017). Optimization strategies in design space exploration. In Handbook of Hardware/Software Codesign (pp. 189-216). Springer Netherlands.

Tundo, A., Mobilio, M., Ilager, S., Brandić, I., Bartocci, E., & Mariani, L. (2023, September). An Energy-Aware Approach to Design Self-Adaptive AI-based Applications on the Edge. In 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE) (pp. 281-293). IEEE.

UNIVERSITA DEGLI STUDI DI MILANO BICOCCA

DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

TU WIEN

FAKULTÄT FÜR !NFORMATIK

Faculty of Informatics

HPC Lab

High Performance Computing Laboratory